

## Durham Research Online

---

### Deposited in DRO:

27 January 2012

### Version of attached file:

Published Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Sun, G. and Holliman, N. S. (2009) 'Evaluating methods for controlling depth perception in stereoscopic cinematography.', in Stereoscopic displays and applications XX. Bellingham, WA: SPIE, 72370I. Proceedings of SPIE. (7237).

### Further information on publisher's website:

<https://doi.org/10.1117/12.807136>

### Publisher's copyright statement:

Copyright 2009 Society of Photo-Optical Instrumentation Engineers. One print or electronic copy may be made for personal use only. Systematic electronic or print reproduction and distribution, duplication of any material in this paper for a fee or for commercial purposes, or modification of the content of the paper are prohibited.

### Additional information:

## Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

# Evaluating Methods for Controlling Depth Perception in Stereoscopic Cinematography

Geng Sun and Nick Holliman

Department of Computer Science, Durham University, United Kingdom

## ABSTRACT

Existing stereoscopic imaging algorithms can create static stereoscopic images with perceived depth control function to ensure a compelling 3D viewing experience without visual discomfort. However, current algorithms do not normally support standard *Cinematic Storytelling* techniques. These techniques, such as object movement, camera motion, and zooming, can result in dynamic scene depth change within and between a series of frames (shots) in stereoscopic cinematography. In this study, we empirically evaluate the following three types of stereoscopic imaging approaches that aim to address this problem.

(1) *Real-Eye Configuration*: set camera separation equal to the nominal human eye interpupillary distance. The perceived depth on the display is identical to the scene depth without any distortion. (2) *Mapping Algorithm*: map the scene depth to a predefined range on the display to avoid excessive perceived depth. A new method that dynamically adjusts the depth mapping from scene space to display space is presented in addition to an existing fixed depth mapping method. (3) *Depth of Field Simulation*: apply Depth of Field (DOF) blur effect to stereoscopic images. Only objects that are inside the DOF are viewed in full sharpness. Objects that are far away from the focus plane are blurred.

We performed a human-based trial using the ITU-R BT.500-11 Recommendation to compare the depth quality of stereoscopic video sequences generated by the above-mentioned imaging methods. Our results indicate that viewers' practical 3D viewing volumes are different for individual stereoscopic displays and viewers can cope with much larger perceived depth range in viewing stereoscopic cinematography in comparison to static stereoscopic images. Our new dynamic depth mapping method does have an advantage over the fixed depth mapping method in controlling stereo depth perception. The DOF blur effect does not provide the expected improvement for perceived depth quality control in 3D cinematography. We anticipate the results will be of particular interest to 3D filmmaking and real time computer games.

**Keywords:** Stereoscopic Cinematography, Depth Perception, Human Factors, Depth of Field, Rendering, 3D Display

## 1. INTRODUCTION

To date, a wide range of research has been devoted to the development of stereoscopic displays and applications. Existing stereoscopic display systems can produce stereoscopic pairs with good image quality (brightness, high resolution, full colour and low crosstalk between left and right images). However, the introduction of stereoscopic displays on public consumer market is yet to be accomplished. The main reason given for this is associated with the issues of visual discomfort caused by the breakdown of Vergence and Accommodation in viewing stereoscopic images.<sup>2-5</sup>

A conclusion has been drawn by all human factors studies:<sup>6-11</sup> viewer's Geometry of Perceived Depth (GPD) on stereoscopic display should be limited into a defined volume, the so-called Comfortable Viewing Range (CVR). When viewing stereoscopic displays the link between eye vergence and accommodation (focus) is thought to be constantly under pressure by requiring the viewer's eyes to verge a long way off the display plane but still focus on the display plane. If the GPD on the display exceeds the limits of CVR, viewers will experience visual discomfort in form of eye strain or headache. In extreme cases, viewers may not even be able to fuse the stereo pairs into one single stereoscopic image.<sup>7</sup>

---

Further author information: Send correspondence to Geng Sun

Geng Sun: E-mail: geng.sun@durham.ac.uk, Telephone: +44 (0) 191 334 1711

## 2. BACKGROUND

### 2.1 3D Movie Renaissance

Since the year 2000, movie production companies and distributors have been showing an increasing interest in stereoscopic format filmmaking. Studios like Disney and Dreamwork are considering creating movies in stereoscopic format natively. Quite a few stereoscopic movies with decent stereoscopic depth quality have already been released, including Intel’s *CyberWorld 3D* in 2000, *Santa vs the Snowman* from Universal Studio in 2002, and *The Polar Express* (2004) and *Beowulf* (2007) from Sony Pictures. About 20 stereoscopic filmmaking projects, such as James Cameron’s *Avatar* and Steven Spielberg’s *Tintin*, are in the works at the time this paper is written.<sup>12</sup>

However, a practical solution for depth mapping from scene space to display space in stereoscopic cinematography is still missing. Instead filmmakers have to manually adjust the camera separation for every frame in the film to ensure the creation of comfortable stereoscopic material.<sup>13–15</sup> This is a very time-consuming and tedious process. In order to overcome this problem, one must start with a basic understanding of Cinematic Storytelling techniques and the new research challenges brought by applying these techniques to stereoscopic cinematography.

### 2.2 Cinematic Storytelling in 3D

The term “Cinematic Storytelling” refers to the non-dialogue techniques, such as object movement, camera motion, zooming, frame composing and editing, used in the movies for conveying ideas.<sup>16</sup> As cinematic storytelling often operates on our subconscious, viewers are more used to certain cinematic patterns which repeatedly appear in movies. For example, objects moving down the screen appear more natural to viewers than objects moving up as it is assumed that they are assisted by gravity. Therefore, if viewers see an object moving up the screen, they will automatically pay more attention to it. This is the reason that the 3D movie can create compelling and enjoyable viewing experience more efficiently by adopting cinematic storytelling techniques.

However, employing cinematic storytelling techniques often involves changing the scene depth that is the volume or boundaries of all the objects inside the scene. The scene depth change makes no difference in traditional 2D movie as everything appears to be flat on the screen. But stereoscopic contents produce screen disparity which changes proportionally with the magnitude of scene depth with a fixed camera separation. Figure 1 demonstrates that the GPD may exceed the limits of CVR with large screen disparity. For this reason considerations of the depth mapping from the scene space to a limited range on the display are required for stereoscopic cinematography.

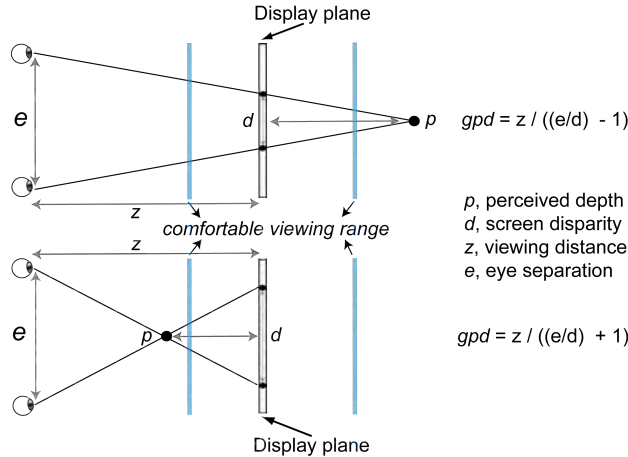


Figure 1. Geometry of Perceived Depth, GPD

### 3. PREVIOUS WORK

Human factors studies had been carried out to define a standard CVR for stereoscopic images. The experiment performed by Yeh and Silverstein<sup>6</sup> suggested that the limits of CVR are 27.0 min arc for crossed disparity which introduces stereoscopic depth in front of screen and 24.0 min arc for uncrossed disparity which introduces stereoscopic depth behind screen. Williams and Parrish<sup>7</sup> concluded that the practical viewing volume for view stereoscopic display should fall between -25% (in front of display surface) and +60% (behind display surface) of the viewer-to-screen distance. They also asserted that the usable viewing volume could be expanded by increasing the viewing distance. Jones *et al*'s experiment<sup>11</sup> showed that the comfortable perceived depth range for desktop displays should be put 50mm in front of and 60mm behind the display surface with a viewing distance of 700 mm.

Jones *et al* developed an algorithm<sup>11</sup> that can map a given scene depth range into any defined single region GPD automatically. The near and far limits of a scene depth were identified and the camera separation was then calculated. It did not require the user to adjust camera separation (multiple times) to achieve the desired GPD and instead it directly provided the exact GPD range specified by users. Holliman further designed a new piecewise, so-called Region of Interest (ROI) algorithm.<sup>19,20</sup> This method was a piecewise approach allowing user to subjectively partition the scene depth volume, with freedom to allocate preferential stereoscopic depth to the region of interest; an approach that can be seen as zooming in depth. Those methods achieved intuitive and precise depth mapping from scene to display space.

Speranza *et al*<sup>21</sup> investigated the relationship between perceived depth, object motion and visual comfort using stereoscopic video sequences. Their results suggested that the speed of perceived depth change might be more important than the absolute magnitude of the perceived depth in determining visual comfort. However, the toed-in camera model used in this study to construct the test stimulus was well known for creating the undesired visual artifact named *Vertical Parallax*.<sup>2,9,13,22,24</sup>

It was Wopking<sup>10</sup> who first proposed the idea of improving the 3D viewing comfort by imitating the DOF effect of natural vision. Blohm *et al*<sup>17</sup> conducted an experiment with eight people, where subjects were asked to watch stereoscopic videos with different volumes of DOF and give their ratings of visual comfort, using the continuous version of the five-grade ITU recommendation impairment scale.<sup>18</sup> Their results indicated that the DOF effect did have a positive effect on improving viewing comfort of stereoscopic video sequences.

### 4. METHODS

We provide a new dynamic depth mapping method that can automatically adjust the camera separation according to the scene depth change in real time so that viewer's GPD stays constant on the display. Together with this new approach, we evaluate four other existing methods to establish a baseline for choosing perceived depth control method in stereoscopic cinematography.

#### 4.1 Real-Eye Configuration

As shown in Figure 2, the camera separation is set equal to the nominal human eye interpupillary distance of 65 mm<sup>23</sup> with the goal to simply map the whole scene on top of the display. We expect the 3D video sequence generated by this method will cause viewing discomfort for viewers as human factors studies have confirmed the need of compressing scene depth around display space in viewing stereoscopic materials. We include this method as the benchmark for comparing different perceived depth control methods.

#### 4.2 Fixed Mapping from Scene Depth to GPD

The fixed mapping method used for this work is adopted from the algorithm designed by Jones *et al*.<sup>11</sup> The camera separation  $A$  is set up using:

$$A = \frac{2Z' \tan \frac{\theta}{2} D_N N'}{W(Z' - N') + D_N N'} \quad (1)$$

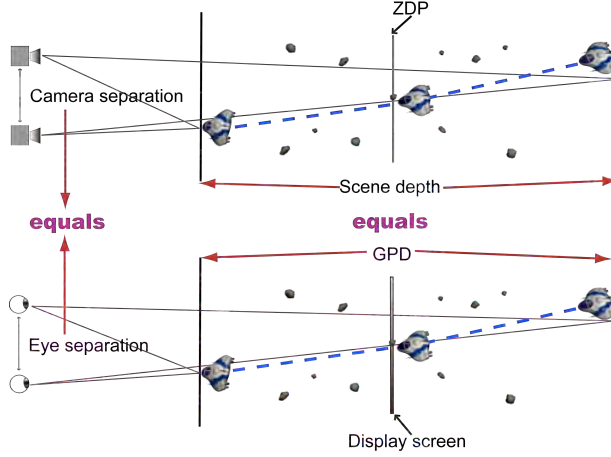


Figure 2. Real-Eye Configuration

$Z'$  is the distance of the cameras from the virtual display (ZDP) in the scene;  $W$  is the width of the physical display screen;  $\theta$  is the field of view of the camera frustum;  $N'$  is distance from the cameras to the closest visible point in the scene;  $D_N$  is the crossed screen disparity. The only unknown variable in (1) is  $D_N$  which can be easily derived using:

$$D_N = \frac{NE}{Z - N} \quad (2)$$

where  $N$  is the closest distance at which objects should appear to the viewer,  $E$  is the human eye separation, and  $Z$  is the viewing distance. All of them can be specified subjectively by the viewer.

This method can guarantee any object in the scene appear inside the CVR and no excessive GPD will be perceived. However, the scene can be very large for real-time graphics despite the fact that sometimes only a small fraction of its volume is actually occupied. Therefore, the perceived depth could be much smaller than the CVR when there is a substantial difference between the limits of the scene boundaries and the actual occupied volume of scene depth, as shown in Figure 3.

### 4.3 Our New Dynamic Mapping from Scene Depth to GPD

This method uses the fixed mapping technique with an additional function that can automatically adjust the camera separation according to the actual occupied scene depth volume in real time so that viewer's GPD will always utilise the whole volume of the CVR, as shown in Figure 4. The real-time update of camera separation can be achieved using the Z-buffer value in OpenGL.

#### 4.3.1 Implementation of Z-Buffer

The Z-Buffer also known as depth buffer in OpenGL stores every pixel's depth information frame by frame in the form of a two-dimensional array (horizontal resolution \* vertical resolution). This array can be accessed through:

```
glEnable(GL_DEPTH_TEST);  
glReadPixels(x, y, width, height, format, type, *pixels);
```

$x, y$  specify the window coordinates of the first pixel that is read from the frame buffer;  $width, height$  specify the dimensions of the pixel rectangle;  $format$  specifies the format of the pixel data;  $type$  specifies the data type of the pixel data and  $pixels$  stores the Z-Buffer value.

In order to convert the Z-Buffer value into the scene depth value, we first need to convert the range of Z-Buffer value from  $[0, 1]$  to  $[-1, 1]$  by:

$$zValue_{new} = zValue_{old} \times 2 - 1 \quad (3)$$

Given (3), the scene depth value can be derived from inverting projection matrix on the Z coordinate:

$$SceneDepthValue = \frac{2 \times farZ \times nearZ}{zValue_{new} \times (farZ - nearZ) - (farZ + nearZ)} \quad (4)$$

$nearZ$ ,  $farZ$  are the distances to the near and far clipping planes.

Replacing  $N'$  in (1) with  $SceneDepthValue$ , the camera separation,  $A$ , can be calculated based on the actual occupied scene depth volume for every frame rendered using the off-screen rendering technique in OpenGL. Although this real-time update of camera separation brings extra computational costs, modern graphics system should still be able to render the scene smoothly.

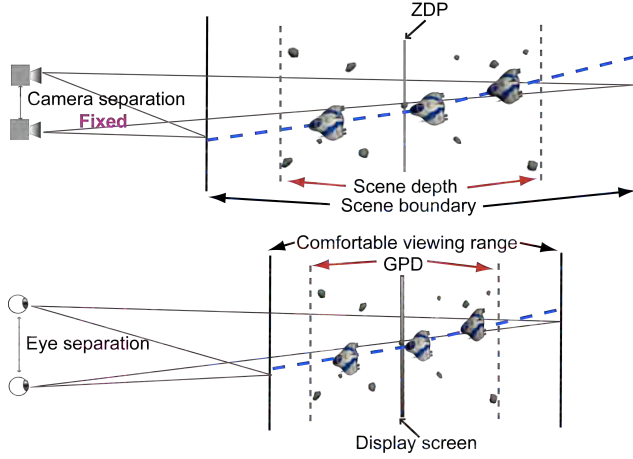


Figure 3. Fixed Mapping Method

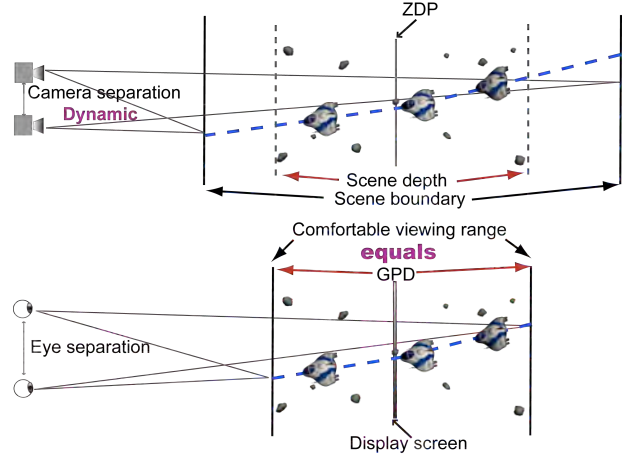


Figure 4. Dynamic Mapping Method

#### 4.4 Fixed Depth of Field Simulation

The Fixed DOF effect is simulated using multi-pass rendering and accumulation buffer in OpenGL. This method, acting like a real camera lens, has a fixed volume of Depth of Field inside which objects appear in full sharpness. Objects that fall out of the DOF are blurred. The level of blurring is proportional to the distance from the object to the focus plane, which is the ZDP in this work as shown in Figure 5. The camera separation used in this approach is also the nominal eye separation, 65 mm.

Literature<sup>10,17</sup> suggested that this method could have the effect of expanding practical 3D viewing volume on the display as objects that could create excessive GPD are blurred, and in turn, no strong stimuli arise for visual system to fuse them.

#### 4.5 Dynamic Depth of Field Simulation

Regarding Figure 6, this method too uses a real-eye camera separation. It has a dynamic focus plane which follows the flight path of the spaceship. As a result, the spaceship, the only moving object in the scene, always stays inside the DOF. The static objects (asteroids) go in and out of focus depending on their distances to the plane of spaceship. We implemented this approach to mimic a natural viewing experience, with the assumption that viewers will spontaneously focus on the moving object in this specific scenario of scene structure.

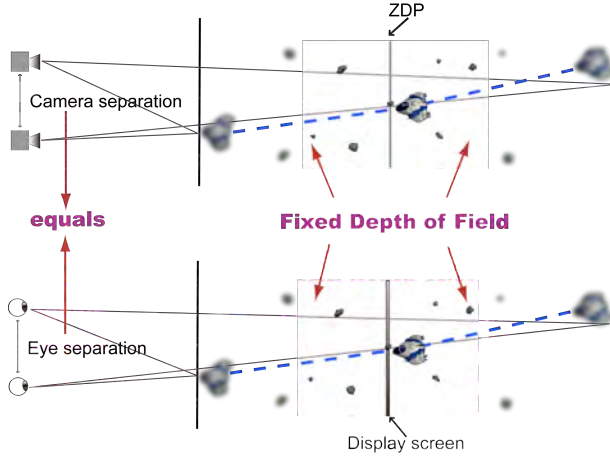


Figure 5. Fixed DOF Simulation

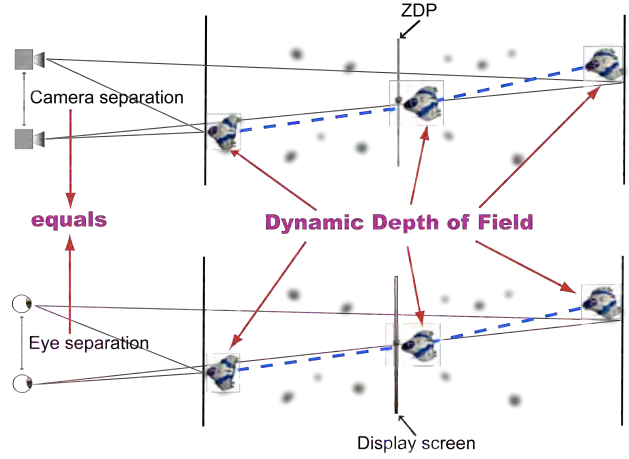


Figure 6. Dynamic DOF Simulation

## 5. EXPERIMENT

### 5.1 Objective

In order to provide a baseline for choosing stereoscopic imaging method to control depth perception in stereoscopic cinematography, we performed a subjective human-based trial, based on the ITU-R BT.500-11 Recommendation,<sup>18</sup> to assess the depth quality of stereoscopic video sequences generated by the five different perceived depth control methods discussed in Section 4.

### 5.2 Hypothesis

Our prediction is that both Mapping algorithms and DOF blur effect simulations will have an advantage over the Real-Eye Configuration method in controlling depth perception in 3D cinematography. We also expect that the video sequence created by our new Dynamic Mapping approach will have a higher depth quality score than the one generated by the Fixed Mapping algorithm. We are not certain about viewer's preference between the Fixed and Dynamic DOF simulations.

### 5.3 Subjects

Seventeen subjects, fifteen male and two female, took part in the experiment. Their ages varied from 20 to 32 with a mean of 24 years. Subjects were not aware of the purpose of the experiment and they were all non-expert, in that their normal work do not concern stereoscopic graphics. All the participants meet the minimum criteria of acuity of 20:30 vision, stereo-acuity at 40 sec-arc and passed the colour vision test. All of them received a nominal payment of five pounds.

### 5.4 Protocol

#### 5.4.1 Procedure

The trial consists of two parts, training session and test session. The trial begins with a training session which demonstrates the range and the type of the scenarios to be assessed. Video sequences played in the training session are different from those played in the test session, but of comparable sensitivity. The results from the training session will not be taken into account in the results analysis.

The procedure of test session follows the Single Stimuli with Multiple Repetition (SSMR) method from the ITU Recommendation. Each test video sequence is played three times organising the test session into three presentations, every presentation includes all five video sequences generated by the five imaging methods only once. Each subject watches the test sequences in a different order which follows the limitations that each test video sequence is not located in the same position in the other presentations and not immediately located before the same sequence in the other presentations. The 3D video sequences will be played on a stereoscopic display



and the viewer will be required to sit down in front of the display at a fixed viewing distance to watch them. An additional 2D display will show the quality scoring sliders and the viewer will be asked to record his/her results via this screen. Each test sequence lasts 20 seconds, followed by a two-second blank interval of gray so that viewers will have time to direct their focus back on the 3D screen before the next video sequence is played. The first presentation is used to stabilise the viewer's opinion. The data issued from this presentation will not be included in the results of the test. The scores assigned to the video sequences are obtained by taking the mean of the data issued from the last two presentations.

All participants will be given the chance to ask questions before, during and after the trial and understand they are free to withdraw from the experiment at any time. Subjects will be fully debriefed when they finishes, regardless of whether they actually complete the experiment. The three vision tests takes about 10 minutes. The training session takes about 5 minutes and the test session lasts about 15 minutes including small breaks after each presentation.

#### 5.4.2 Scoring Scale

Figure 8 illustrates the scoring scale showing on the 2D screen. In each trial the video sequences are rated on a sliding scale of Excellent, Good, Fair, Poor, and Bad. These terms categorise the five different levels and they are the same as those normally used in the ITU-R recommendation. The terms are associated with the value intervals of 100 to 80, 80 to 60, 60 to 40, 40 to 20, and 20 to 0, respectively, providing a continuous rating system. Subjects are asked to score the depth quality of each stereoscopic video sequence by moving the slide bar to the desired position along the scale. The vertical scale displayed on the 2D screen is ten centimetres long and divided into five equal lengths. Results are recorded once subjects click the "Submit Scores" button.

#### 5.5 Apparatus and Viewing Conditions

The experiment was run by a Dell Precision PWS670 computer with Intel Xeon CPU of 3.00GHz 2.99 GHz, 4.00 GB RAM and NVIDIA Quadro FX5600 graphics card. A 24-inch True 3Di stereoscopic display with a 1920 × 1200 resolution, as shown in Figure 7, was used to play the experimental test video sequences. This 3D display requires viewers to wear polarised glasses to fuse the left and right images. The scoring scale was shown on a 21-inch HITACHI CRT with a resolution of 1280×1024. We chose a CRT monitor for displaying the scoring scale so that viewers did not need to take on/off glasses when switching between 3D and 2D displays.

The two displays run independently. However, both of them used the graphics card from NVIDIA Quadro FX family and were driven by NVIDIA ForceWare Release 80. Viewers were asked to sit exactly 700 mm in front of the 3D display. The whole experiment was conducted in a dimly lit room.



Figure 7. Displays



Figure 8. Scoring Scale



## 5.6 Stimulus

The five different stimulus tested in the experiment are shown in figure 10, 11, 12, 13 and 14. They all have the same frame composition. The Computer Graphics (CG) animated scene is composed of a flying spaceship and eleven still asteroids (one of them is occluded by the spaceship in the figures). The spaceship is flying back and forth through those asteroids along a curvilinear path resembling a figure of eight. Its velocity varies smoothly along the flight trajectory. The spaceship will slow down as it turns around to avoid any undesired visual artifacts that could be caused by sharp and sudden turns. Figure 13 demonstrates the Fixed DOF stimuli. The spaceship is heavily blurred as it is quite far away from the focus plane on which the central asteroid is located. Figure 14 illustrates the Dynamic DOF stimuli. We can see that the asteroid at bottom-right corner is also inside the DOF and seen in full sharpness due to the close distance between itself and the plane of the spaceship.

3D Studio Max was used to model the CG animated scene, edit the flight path of the spaceship and generate its coordinates. Our software read in the exported scene objects and flight path coordinates and rendered the scene frame by frame in OpenGL. Each test stimuli had 500 frames in total with a frame rate of 25 fps. Viewer's maximum GPD of Real-Eye Configuration, Fixed DOF and Dynamic DOF stimulus was 200 mm in front of the display screen and 240 mm behind the display screen, which was a 1:1 mapping of the scene depth maximum and four times as large as the CVR defined by Jones *et al.* Viewer's maximum GPD of Fixed and Dynamic Mapping approaches followed Jones *et al.*'s recommendation, which was 50 mm and 60 mm on each side of the display.

## 6. RESULTS AND ANALYSIS

Figure 9 is the box plot of the results from all seventeen subjects. The mean score and standard deviation for each method is shown in Table 1. The results from one-way Single Factor ANOVA indicated that the differences between different methods were statistically significant,  $F$  ratio = 19.117 >  $F$  critical = 2.486. A Paired T-Test was then performed on every possible interaction. The results of the T-Test comparisons concerning the Real-Eye Configuration is listed in Table 2.

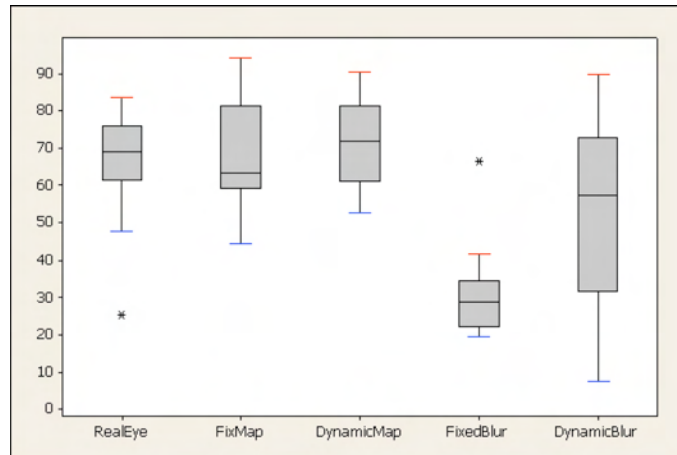


Figure 9. Box Plot Results

As shown in Table 1, the mean scores of Real-Eye, Fixed Mapping and Dynamic Mapping methods ( $\mu_{re}$ ,  $\mu_{fm}$ ,  $\mu_{dm}$ ) were quite similar. Paired T-Tests were carried out and only difference between Fixed Mapping and Dynamic Mapping approaches was statistically significant,  $H_0 : \mu_{fm} = \mu_{dm}$  vs.  $H_1 : \mu_{fm} < \mu_{dm}$ ,  $p$  value = 0.028. Neither Dynamic mapping nor Fixed mapping was able to provide an advantage over the Real-Eye approach as expected.

Only two subjects out of seventeen reported that they did not like the stereoscopic videos generated by Real-Eye method as they provided too much perceived depth. All other subjects had no problem fusing the stereo pairs with perceived depth that was four times as large as the CVR suggested by Jones *et al.* These results suggested that the CVR defined by Jones *et al.* was too conservative for viewing stereoscopic video sequences on the test display.

Table 1. Mean and Standard Deviation

Method	Mean, $\mu$	St. Dev.
Real-Eye	$\mu_{re} = 67.088$	13.989
FixedMap	$\mu_{fm} = 68.441$	14.414
DynamicMap	$\mu_{dm} = 71.206$	11.214
FixedBlur	$\mu_{fb} = 31$	11.413
DynamicBlur	$\mu_{db} = 52.882$	24.223

Table 2. T-Test results concerning the Real-Eye method

Hypotheses	$p$ value	Conclusion
$H_0 : \mu_{re} = \mu_{fm}$ vs. $H_1 : \mu_{re} < \mu_{fm}$	0.401	Fail to reject $H_0$
$H_0 : \mu_{re} = \mu_{dm}$ vs. $H_1 : \mu_{re} < \mu_{dm}$	0.206	Fail to reject $H_0$
$H_0 : \mu_{re} = \mu_{fb}$ vs. $H_1 : \mu_{re} < \mu_{fb}$	1.000	Fail to reject $H_0$
$H_0 : \mu_{re} = \mu_{db}$ vs. $H_1 : \mu_{re} < \mu_{db}$	0.985	Fail to reject $H_0$

As shown in Table 1, only the mean ( $\mu_{fb}$ ) of Fixed DOF method fell below 50 which corresponded with the term “Fair” in ITU’s grading scale. One subject out of seventeen rated the Fixed DOF method above “Fair”. The other sixteen all stated that they did not like this approach as they would like to see the whole context when viewing 3D videos. Two of them said that they did not mind blurring the background objects much, it was the foreground blur that really annoyed them. Although the mean,  $\mu_{db}$ , was a little bit higher than “Fair”, the Dynamic DOF method had the biggest standard deviation. Scores assigned by subjects varied from 7.5, the lowest score for all methods, to 90. People disliked it for the same reason as they disliked the Fixed DOF. Those who really liked it expressed that they spontaneously focused on the flying spaceship and believed this method was a good simulation of natural vision (we did not ask viewers to specifically focus on the spaceship). The results from a Paired T-Test comparison between Fixed and Dynamic DOF methods revealed that there was a 100% probability that the difference between Dynamic DOF simulation and Fixed DOF simulation was statistically significant,  $H_0 : \mu_{fb} = \mu_{db}$  vs.  $H_1 : \mu_{fb} < \mu_{db}$ ,  $p$  value = 0.000.

Table 3. T-Test results between mapping methods and DOF simulations

Hypotheses	$p$ value	Conclusion
$H_0 : \mu_{fm} = \mu_{dm}$ vs. $H_1 : \mu_{fm} < \mu_{dm}$	0.028	Reject $H_0$
$H_0 : \mu_{fb} = \mu_{db}$ vs. $H_1 : \mu_{fb} < \mu_{db}$	0.000	Reject $H_0$
$H_0 : \mu_{fb} = \mu_{fm}$ vs. $H_1 : \mu_{fb} < \mu_{fm}$	0.000	Reject $H_0$
$H_0 : \mu_{fb} = \mu_{dm}$ vs. $H_1 : \mu_{fb} < \mu_{dm}$	0.000	Reject $H_0$
$H_0 : \mu_{db} = \mu_{fm}$ vs. $H_1 : \mu_{db} < \mu_{fm}$	0.019	Reject $H_0$
$H_0 : \mu_{db} = \mu_{dm}$ vs. $H_1 : \mu_{db} < \mu_{dm}$	0.000	Reject $H_0$

Paired T-Tests were also performed between the two depth mapping methods and the two DOF simulations. The results, listed in Table 3, indicated that differences between depth mapping methods and DOF simulations in general were significant.

## 7. CONCLUSION AND DISCUSSION

We performed a subjective human-based experiment to evaluate different stereoscopic imaging algorithms that aim to better control the perceived depth in stereoscopic cinematography. In our experiment, viewers can cope with perceived depth that is four times as large as the range defined by Jones *et al*'s study. This result suggested that the practical viewing volume on 3D display differs between individual displays and viewers' comfortable 3D viewing ranges are expanded in viewing stereoscopic cinematography in comparison to static stereoscopic images.

Our new approach of dynamic mapping of depth from the scene space to display space scored the highest mean among all the tested approaches. Statistics confirmed that it was able to provide a significant effect over the fixed mapping algorithm and DOF simulations in controlling the perceived depth in stereoscopic cinematography.

We also learned that, in contrast to the conclusions drawn by previous studies, the Depth of Field blur simulation does not improve the perceived depth quality in 3D cinematography. However, there were indications in our results suggesting that viewers could regard the Dynamic DOF simulation as a good imitation of natural visual experience when there are both dynamic and static objects in the scene.

## 8. FUTURE WORK

Future work will mainly focus on: 1) define practical comfortable viewing ranges for stereoscopic cinematography on different stereoscopic displays. 2) investigate more advanced DOF simulation techniques, e.g., ray distribution approach, to define the optimum DOF for 3D image representation. 3) design and evaluate perceived depth control methods for other cinematic storytelling techniques, such as camera motion, zooming, etc.

## 9. ACKNOWLEDGEMENTS

The authors gratefully acknowledge their colleague, Mr. Paul Gorley, for providing the 3D video testing software used in the trial. Thanks are also due to the Department of Computer Science at Durham University for purchasing the stereoscopic displays used in this work.

## REFERENCES

- [1] Ohm, J. R., "Stereo/multiview video encoding using the MPEG family of standards," in [*Stereoscopic Displays and Virtual Reality Systems VI. Proceedings of the SPIE*], **3639**, 242–253 (May 1999).
- [2] McAllister, D. F., ed., [*Stereo computer graphics: and other true 3D technologies*], Princeton University Press, Princeton, NJ, USA (1993).
- [3] Sugihara, T., Miyasato, T., and Nakatsu, R., "An evaluation of visual fatigue in 3-d displays: focusing on the mismatching of convergence and accommodation," *IEICE TRANS. ELECTRON.*, **E82-C**(10), 1814–1822 (1999).
- [4] Lambooi, M. T. M., IJsselstein, W. A., and Heynderickx, I., "Visual discomfort in stereoscopic displays: a review," in [*Stereoscopic Displays and Virtual Reality Systems XIV. Proceedings of the SPIE*], Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference **6490** (Feb. 2007).
- [5] Hoffman, D. M., Girshick, A. R., Akeley, K., and Banks, M. S., "Vergence-accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of vision* **8**(3) (2008).
- [6] Yeh, Y. Y. and Silverstein, L. D., "Limits of fusion and depth judgment in stereoscopic color displays," *Hum. Factors* **32**(1), 45–60 (1990).
- [7] Williams, S. P. and Parrish, R. V., "New computational control techniques and increased understanding for stereo 3d displays," in [*Proceedings of SPIE Stereoscopic Displays and Applications*], 73–82 (1990).
- [8] Hiruma, N. and Fukuda, T., "Accommodation response to binocular stereoscopic tv images and their viewing conditions," *SMPTE Journal*, 1137–1144 (Dec. 1993).
- [9] Woods, A. J., Docherty, T., and Koch, R., "Image distortions in stereoscopic video systems," in [*Stereoscopic Displays and Applications IV. Proceedings of the SPIE*], **1915**, 36–47 (Sept. 1993).
- [10] Wopking, M., "Viewing comfort with stereoscopic pictures: An experimental study on the subjective effects of disparity magnitude and depth of focus," *Journal of the SID* **3**(3), 101–103 (1995).
- [11] Jones, G. R., Lee, D., Holliman, N. S., and Ezra, D., "Controlling perceived depth in stereoscopic images," in [*Stereoscopic Displays and Virtual Reality Systems VIII. Proceedings of the SPIE*], **4297**, 42–53 (June 2001).
- [12] Autodesk, "Stereoscopic filmmaking whitepaper." [http://images.autodesk.com/adsk/files/stereoscopic\\_whitepaper\\_final108.pdf](http://images.autodesk.com/adsk/files/stereoscopic_whitepaper_final108.pdf).
- [13] Holliman, N., Baugh, C., Frenk, C., Jenkins, A., Froner, B., Hassaine, D., Helly, J., Metcalfe, N., and Okamoto, T., "Cosmic cookery: making a stereoscopic 3D animated movie," in [*Stereoscopic Displays and Virtual Reality Systems XIII. Proceedings of the SPIE*], **6055**, 34–45 (Feb. 2006).
- [14] Murray, H., "3D animation in three dimensions: the rocky road to the obvious," in [*Stereoscopic Displays and Applications XIII. Proceedings of the SPIE*], **6803** (Feb. 2006).
- [15] Engle, R., "Beowulf 3D: A Case Study," in [*Stereoscopic Displays and Applications XIX. Proceeding of the SPIE*], **6055** (2008).

- [16] Sijll, J. V., [*Cinematic Storytelling: The 100 Most Powerful Film Conventions Every Filmmaker Must Know*], Michael Wiese Productions, Cambridge, Massachusetts (2005).
- [17] Blohm, W., Beldie, I. P., Schenke, K., Fazel, K., and Pastoor, S., "Stereoscopic image representation with synthetic depth of field," *Journal of the Society for Information Display* **5**(3), 307–313 (1997).
- [18] Union, I. T., "Methodology for the subjective assessment of the quality of television pictures," (2002). <http://www.itu.int/rec/R-REC-BT.500-11-200206-I>.
- [19] Holliman, N. S., "Mapping perceived depth to regions of interest in stereoscopic images," in [*Stereoscopic Displays and Virtual Reality Systems XI. Proceedings of the SPIE*], **5291**, 117–128 (May 2004).
- [20] Holliman, N. S., "Smoothing region boundaries in variable depth mapping for real-time stereoscopic images," in [*Stereoscopic Displays and Virtual Reality Systems XII. Proceedings of the SPIE*], **5664**, 281–292 (Mar. 2005).
- [21] Speranza, F., Tam, W. J., Renaud, R., and Hur, N., "Effect of disparity and motion on visual comfort of stereoscopic images," in [*Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*], *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* **6055**, 94–103 (Feb. 2006).
- [22] Kooi, F. L. and Toet, A., "Visual comfort of binocular and 3d displays," *Displays* **25**, 99–108 (August 2004).
- [23] Dodgson, N. A., "Variation and extrema of human interpupillary distance," in [*Stereoscopic Displays and Virtual Reality Systems XI. Proceedings of the SPIE*], Woods, A. J., Merritt, J. O., Benton, S. A., and Bolas, M. T., eds., *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference* **5291**, 36–46 (May 2004).
- [24] Ware, C., Gobrecht, C., and Paton, M. A., "Dynamic adjustment of stereo display parameters," *IEEE Transactions on Systems, Man and Cybernetics* **28**, 56–65 (January 1998).



Figure 10. Real-Eye Stimuli



Figure 11. Fixed Mapping Stimuli



Figure 12. Dynamic Mapping Stimuli

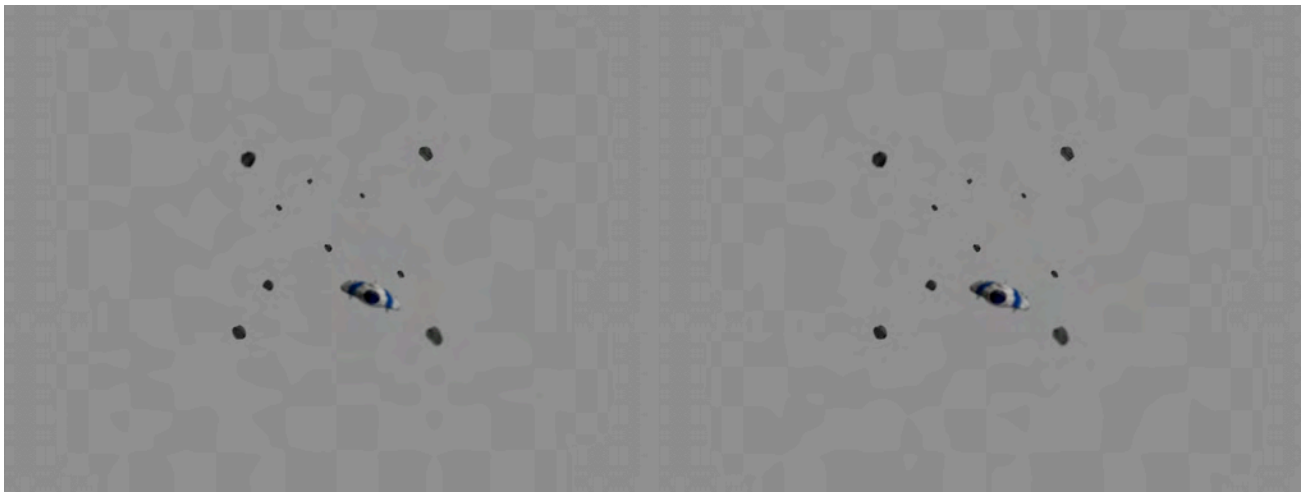


Figure 13. Fixed DOF Stimuli

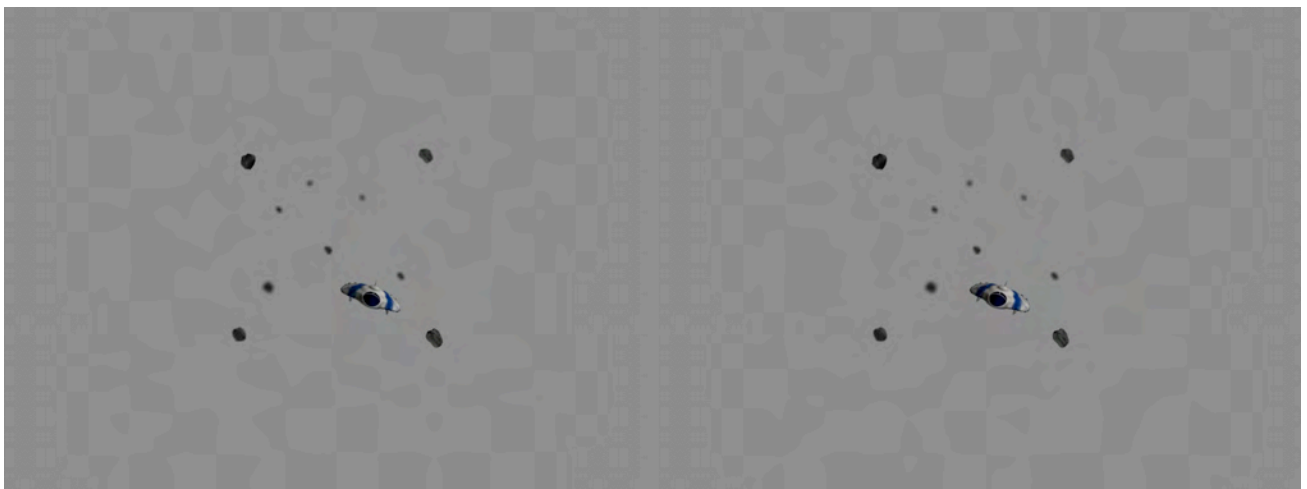


Figure 14. Dynamic DOF Stimuli